```
RAW SEQUENCE LISTING              DATE: 11/01/2001
PATENT APPLICATION:  US/09/845,226    TIME: 13:46:02

       Input Set : N:\Crf3\RULE60\09845226.txt
       Output Set: N:\CRF3\11012001\I845226.raw
```

```
 3 <110> APPLICANT: Tang, Jordan J.N.
 4         Hong, Lin
 5         Ghosh, Arun K.
 7 <120> TITLE OF INVENTION: Inhibitors of Memapsin 2 and Use Thereof
 9 <130> FILE REFERENCE: OMRF 182
11 <140> CURRENT APPLICATION NUMBER: 09/845,226
12 <141> CURRENT FILING DATE: 2001-04-30
14 <150> PRIOR APPLICATION NUMBER: 09/603,713
15 <151> PRIOR FILING DATE: 2000-06-27
17 <150> PRIOR APPLICATION NUMBER: 60/168,060
18 <151> PRIOR FILING DATE: 1999-11-30
20 <150> PRIOR APPLICATION NUMBER: 60/177,836
21 <151> PRIOR FILING DATE: 2000-01-25
23 <150> PRIOR APPLICATION NUMBER: 60/178,368
24 <151> PRIOR FILING DATE: 2000-01-27
26 <150> PRIOR APPLICATION NUMBER: 60/210,292
27 <151> PRIOR FILING DATE: 2000-06-08
29 <160> NUMBER OF SEQ ID NOS: 31
31 <170> SOFTWARE: PatentIn Ver. 2.1
33 <210> SEQ ID NO: 1
34 <211> LENGTH: 3252
35 <212> TYPE: DNA
36 <213> ORGANISM: Homo sapiens
38 <400> SEQUENCE: 1
39 gcgggagtgc tgcctgccca cggcacccag cacggcatcc ggctgcccct gcgcagcggc 60
40 ctgggggcg cccccctggg gctgcggctg ccccgggaga ccgacgaaga gcccgaggag 120
41 cccggccgga ggggcagctt tgtggagatg gtggacaacc tgaggggcaa gtcggggcag 180
42 ggctactacg tggagatgac cgtgggcagc cccccgcaga cgctcaacat cctggtggat 240
43 acaggcagca gtaactttgc agtgggtgct gcccccacc ccttcctgca tcgctactac 300
44 cagaggcagc tgtccagcac ataccgggac ctccggaagg gtgtgtatgt gccctacacc 360
45 cagggcaagt gggaagggga gctgggcacc gacctggtaa gcatcccca tggccccaac 420
46 gtcactgtgc gtgccaacat tgctgccatc actgaatcag acaagttctt catcaacggc 480
47 tccaactggg aaggcatcct ggggctggcc tatgctgaga ttgccaggcc tgacgactcc 540
48 ctggagcctt tctttgactc tctggtaaag cagacccacg ttcccaacct cttctccctg 600
49 cagctttgtg gtgctggctt cccccctcaac cagtctgaag tgctggcctc tgtcggaggg 660
50 agcatgatca ttggaggtat cgaccactcg ctgtacacag gcagtctctg gtatacaccc 720
51 atccggcggg agtggtatta tgaggtgatc attgtgcggg tggagatcaa tggacaggat 780
52 ctgaaaatgg actgcaagga gtacaactat gacaagagca ttgtggacag tggcaccacc 840
53 aaccttcgtt tgcccaagaa agtgtttgaa gctgcagtca atccatcaa ggcagcctcc 900
54 tccacggaga agttccctga tggtttctgg ctaggagagc agctggtgtg ctggcaagca 960
55 ggcaccaccc cttggaacat tttcccagtc atctcactct acctaatggg tgaggttacc 1020
56 aaccagtcct ccgcatcac catccttccg cagcaatacc tgcggccagt ggaagatgtg 1080
57 gccacgtccc aagacgactg ttacaagttt gccatctcac agtcatccac gggcactgtt 1140
58 atgggagctg ttatcatgga gggcttctac gttgtctttg atcgggcccg aaaacgaatt 1200
59 ggctttgctg tcagcgcttg ccatgtgcac gatgagttca ggacggcagc ggtggaaggc 1260
60 ccttttgtca ccttggacat ggaagactgt ggctacaaca ttccacagac agatgagtca 1320
61 accctcatga ccatagccta tgtcatggct gccatctgcg ccctcttcat gctgccactc 1380
```

```
62 tgcctcatgg tgtgtcagtg gcgctgcctc cgctgcctgc gccagcagca tgatgacttt 1440
63 gctgatgaca tctccctgct gaagtgagga ggcccatggg cagaagatag agattcccct 1500
64 ggaccacacc tccgtggttc actttggtca caagtaggag acacagatgg cacctgtggc 1560
65 cagagcacct caggaccctc cccacccacc aaatgcctct gccttgatgg agaaggaaaa 1620
66 ggctggcaag gtgggttcca gggactgtac ctgtaggaaa cagaaaagag aagaaagaag 1680
67 cactctgctg gcgggaatac tcttggtcac ctcaaattta agtcgggaaa ttctgctgct 1740
68 tgaaacttca gccctgaacc tttgtccacc attcctttaa attctccaac ccaaagtatt 1800
69 cttcttttct tagtttcaga agtactggca tcacacgcag gttaccttgg cgtgtgtccc 1860
70 tgtggtaccc tggcagagaa gagaccaagc ttgtttccct gctggccaaa gtcagtagga 1920
71 gaggatgcac agtttgctat ttgctttaga gacagggact gtataaacaa gcctaacatt 1980
72 ggtgcaaaga ttgcctcttg aattaaaaaa aaactagatt gactatttat acaaatgggg 2040
73 gcggctggaa agaggagaag gagagggagt acaaagacag ggaatagtgg gatcaaagct 2100
74 aggaaaggca gaaacacaac cactcaccag tcctagtttt agacctcatc tccaagatag 2160
75 catcccatct cagaagatgg gtgttgtttt caatgttttc ttttctgtgg ttgcagcctg 2220
76 accaaaagtg agatgggaag ggcttatcta gccaaagagc tcttttttag ctctcttaaa 2280
77 tgaagtgccc actaagaagt tccacttaac acatgaattt ctgccatatt aatttcattg 2340
78 tctctatctg aaccaccctt tattctacat atgataggca gcactgaaat atcctaaccc 2400
79 cctaagctcc aggtgccctg tgggagagca actggactat agcagggctg ggctctgtct 2460
80 tcctggtcat aggctcactc tttcccccaa atcttcctct ggagctttgc agccaaggtg 2520
81 ctaaaaggaa taggtaggag acctcttcta tctaatcctt aaaagcataa tgttgaacat 2580
82 tcattcaaca gctgatgccc tataaccct gcctggattt cttcctatta ggctataaga 2640
83 agtagcaaga tctttacata attcagagtg gtttcattgc cttcctaccc tctctaatgg 2700
84 cccctccatt tatttgacta aagcatcrca cagtggcact agcattatac caagagtatg 2760
85 agaaatacag tgctttatgg ctctaacatt actgccttca gtatcaaggc tgcctggaga 2820
86 aaggatggca gcctcagggc ttccttatgt cctccaccac aagagctcct tgatgaaggt 2880
87 catcttttc ccctatcctg ttcttcccct ccccgctcct aatggtacgt gggtacccag 2940
88 gctggttctt gggctaggta gtggggacca agttcattac ctccctatca gttctagcat 3000
89 agtaaactac ggtaccagtg ttagtgggaa gagctgggtt ttcctagtat acccactgca 3060
90 tcctactcct acctggtcaa cccgctgctt ccaggtatgg gacctgctaa gtgtggaatt 3120
91 acctgataag ggagagggaa atacaaggag ggcctctggt gttcctggcc tcagccagct 3180
92 gcccmcaagc cataaaccaa taaamcaaga atactgagtc taaaaaaaaa aaaaaaaaaa 3240
93 aaaaaaaaaa aa                                                      3252
96 <210> SEQ ID NO: 2
97 <211> LENGTH: 488
98 <212> TYPE: PRT
99 <213> ORGANISM: Homo sapiens
101 <220> FEATURE:
102 <223> OTHER INFORMATION: Purified Memapsin 2
104 <220> FEATURE:
105 <223> OTHER INFORMATION: Amino Acids 28-48 are remnant putative propeptide
106        residues
108 <220> FEATURE:
109 <223> OTHER INFORMATION: Amino Acids 58-61, 78, 80, 82-83, 116, 118-121,
110        156, 166, 174, 246, 274, 276, 278-281, 283, and
111        376-377 are residues in contact with the OM99-2
112        inhibitor
114 <220> FEATURE:
115 <223> OTHER INFORMATION: Amino acids 54-57, 61-68, 73-80, 86-89, 109-111,
116        113-118, 123-134, 143-154, 165-168, 198-202, and
```

```
117      220-224 are N-lobe Beta Strands
119 <220> FEATURE:
120 <223> OTHER INFORMATION: Amino Acids 184-191 and 210-217 are N-lobe Helices
122 <220> FEATURE:
123 <223> OTHER INFORMATION: Amino acids 237-240, 247-249, 251-256, 259-260,
124      273-275, 282-285, 316-318, 331-336, 342-348,
125      354-357, 366-370, 372-375, 380-383, 390-395,
126      400-405, and 418-420 are C-lobe Beta Strands
128 <220> FEATURE:
129 <223> OTHER INFORMATION: Amino Acids 286-299, 307-310, 350-353, 384-387,
130      and 427-431 are C-lobe Helices
132 <400> SEQUENCE: 2
133 Ala Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro
134   1           5                  10                  15
136 Leu Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg
137             20                  25                  30
139 Glu Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val
140         35                  40                  45
142 Glu Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val
143       50                  55                  60
145 Glu Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp
146   65                  70                  75                  80
148 Thr Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu
149             85                  90                  95
151 His Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg
152             100                 105                 110
154 Lys Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu
155             115                 120                 125
157 Gly Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg
158       130                 135                 140
160 Ala Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly
161 145                 150                 155                 160
163 Ser Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg
164             165                 170                 175
166 Pro Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr
167             180                 185                 190
169 His Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro
170             195                 200                 205
172 Leu Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile
173       210                 215                 220
175 Gly Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro
176 225                 230                 235                 240
178 Ile Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile
179             245                 250                 255
181 Asn Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys
182             260                 265                 270
184 Ser Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val
185       275                 280                 285
187 Phe Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys
```

RAW SEQUENCE LISTING   DATE: 11/01/2001
PATENT APPLICATION: US/09/845,226   TIME: 13:46:02

Input Set : N:\Crf3\RULE60\09845226.txt
Output Set: N:\CRF3\11012001\I845226.raw

```
188      290                  295                  300
190 Phe Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala
191 305                  310                  315                  320
193 Gly Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met
194                  325                  330                  335
196 Gly Glu Val Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln
197                  340                  345                  350
199 Tyr Leu Arg Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr
200          355                  360                  365
202 Lys Phe Ala Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val
203      370                  375                  380
205 Ile Met Glu Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile
206 385                  390                  395                  400
208 Gly Phe Ala Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala
209                  405                  410                  415
211 Ala Val Glu Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr
212          420                  425                  430
214 Asn Ile Pro Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val
215          435                  440                  445
217 Met Ala Ala Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val
218      450                  455                  460
220 Cys Gln Trp Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe
221 465                  470                  475                  480
223 Ala Asp Asp Ile Ser Leu Leu Lys
224                  485
227 <210> SEQ ID NO: 3
228 <211> LENGTH: 503
229 <212> TYPE: PRT
230 <213> ORGANISM: Homo sapiens
232 <220> FEATURE:
233 <223> OTHER INFORMATION: Pro-memapsin 2
235 <220> FEATURE:
236 <223> OTHER INFORMATION: Amino Acids 1-15 are vector-derived residues
238 <220> FEATURE:
239 <223> OTHER INFORMATION: Amino Acids 16-64 are a putative pro peptide
241 <220> FEATURE:
242 <223> OTHER INFORMATION: Amino Acids 1-13 are the T7 promoter
244 <220> FEATURE:
245 <223> OTHER INFORMATION: Amino Acids 16-456 are Pro-memapsin 2-T1
247 <220> FEATURE:
248 <223> OTHER INFORMATION: Amino Acids 16-421 are Promemapsin 2-T2
250 <400> SEQUENCE: 3
251 Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Gly Ser Met Ala
252  1               5                  10                  15
254 Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro Leu
255                  20                  25                  30
257 Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg Glu
258          35                  40                  45
260 Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val Glu
```

RAW SEQUENCE LISTING   DATE: 11/01/2001
PATENT APPLICATION: US/09/845,226   TIME: 13:46:02

Input Set : N:\Crf3\RULE60\09845226.txt
Output Set: N:\CRF3\11012001\I845226.raw

```
261       50                    55                    60
263 Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val Glu
264  65                    70                    75                    80
266 Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp Thr
267                    85                    90                    95
269 Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu His
270                   100                   105                   110
272 Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg Lys
273            115                   120                   125
275 Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly
276        130                   135                   140
278 Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg Ala
279 145                   150                   155                  160
281 Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser
282                   165                   170                   175
284 Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro
285                   180                   185                   190
287 Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His
288            195                   200                   205
290 Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu
291        210                   215                   220
293 Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly
294 225                   230                   235                   240
296 Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile
297            245                   250                   255
299 Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn
300            260                   265                   270
302 Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser
303        275                   280                   285
305 Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe
306        290                   295                   300
308 Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe
309 305                   310                   315                   320
311 Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly
312            325                   330                   335
314 Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly
315            340                   345                   350
317 Glu Val Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr
318        355                   360                   365
320 Leu Arg Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys
321        370                   375                   380
323 Phe Ala Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val Ile
324 385                   390                   395                   400
326 Met Glu Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile Gly
327            405                   410                   415
329 Phe Ala Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala Ala
330            420                   425                   430
332 Val Glu Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr Asn
333            435                   440                   445
```

VERIFICATION SUMMARY   DATE: 11/01/2001
PATENT APPLICATION: US/09/845,226   TIME: 13:46:03

Input Set : N:\Crf3\RULE60\09845226.txt
Output Set: N:\CRF3\11012001\I845226.raw

L:580 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:22
L:580 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:22
L:580 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22
L:597 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:23
L:597 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:23
L:597 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23